International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Multimodal Secure Communication and Filtering Engine

Sreeja KS[1], Jishnu S[2], Dr. Smita C Thomas[3]

[1]Department of Computer Science and Engineering, Mountzion College of Engineering, Kadammannitta, Pathanamthitta, Kerala, India
Email: *sreeja85[at]gmail.com*

[2]Assistant Professor, Department of Computer Science and Engineering, Mountzion College of Engineering, Kadammannitta, Pathanamthitta, Kerala, India
Email: *isjishnus[at]gmail.com*

[3]Department of Computer Science and Engineering, Mountzion College of Engineering, Kadammannitta, Pathanamthitta, Kerala, India
Email: *easyktulectures[at]gmail.com*

Abstract: ***"Multimodal Secure Communication and Filtering Engine" is designed to deliver a robust and user-friendly communication platform by integrating multiple advanced technologies. The system employs hybrid cryptography to provide strong end-to-end encryption, ensuring that all messages remain confidential and tamper-proof. To enhance user safety, the platform incorporates intelligent content monitoring. Malicious URL detection powered by a Random Forest classifier and hate speech detection using NLP and BiLSTM models work together to identify and block harmful content in real time. This proactive filtering protects users from threats such as phishing, cyber attacks, and abusive language. Security is further reinforced through multi-factor authentication (MFA) options. Users can choose between Basic, Two-Factor, or Three-Factor authentication based on their security needs. The addition of face recognition provides a strong biometric layer, preventing unauthorized access and strengthening overall account protection. By combining encryption, intelligent content filtering, and flexible authentication methods, the system ensures a safer and more reliable online experience for all users.***

Keywords: BiLSTM, NLP, CNN, Random Forest

1. Introduction

The "Multimodal Secure Communication and Filtering Engine" is an innovative and sophisticated project that aims to revolutionize digital communication by prioritizing security and content filtering. In today's interconnected world, the internet and various digital platforms have become integral to communication, making it crucial to ensure privacy and protect users from harmful content and cyber security threats. This project integrates a range of cutting-edge technologies, including cryptography, malicious URL detection using Random Forest algorithm, and hate speech detection using Natural Language Processing (NLP) and BiLSTM (Bidirectional Long Short-Term Memory) models. Additionally, it utilizes face recognition for multi-factor authentication, creating a secure and user-friendly communication system.

With the rising prevalence of cyber-attacks, hate speech, and inappropriate content on digital platforms, traditional communication systems are increasingly vulnerable. The system sets itself apart by leveraging state-of-the-art cryptographic techniques to ensure that all forms of communication are encrypted and secure. By adopting a hybrid cryptography approach combining AES and ECC, the system ensures end-to-end encryption, safeguarding the confidentiality and integrity of user data. Moreover, the project implements a robust content filtering system utilizing advanced machine learning algorithms. The malicious URL detection employs the Random Forest algorithm, which is known for its ability to classify and detect malicious URLs effectively. In contrast, the hate speech detection leverages the power of NLP techniques and BiLSTM models, allowing the system to analyze and identify hate speech content with high accuracy.

When users send messages or images, the system analyzes the content in real-time, proactively filtering out harmful content such as hate speech, explicit images, and malicious URLs. This proactive content filtering ensures that harmful material is intercepted and blocked before it reaches the intended recipient, promoting a safer and more respectful online environment.

To augment the security measures, the project implements a multi-factor authentication system. Users can choose from three authentication levels: Basic, Two Factor, and Three Factor, depending on their desired level of security. Basic authentication requires a standard username and password, while Two Factor authentication adds an extra layer of security through one-time passwords (OTP). The most robust security is achieved with Three Factor authentication, where users undergo face recognition in addition to username, password, and OTP verification.

2. Literature Survey

The paper **Hate Speech Detection using NLP and Deep Learning.** Hate speech detection has gained significant attention due to the increasing misuse of online platforms. Traditional machine learning techniques such as Support Vector Machines (SVM) and Na¨ıve Bayes have been used for text classification; however, they often fail to capture contextual meaning. Studies emphasize the use of deep learning models, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks,

for improved performance. The Bidirectional LSTM (BiLSTM) model has shown higher accuracy as it processes text in both forward and backward directions, capturing contextual dependencies effectively. Additionally, Natural Language Processing (NLP) techniques such as tokenization, stop-word removal, and word embeddings further enhance classification performance.

The paper **Secure Communication using Cryptography.** Secure communication systems primarily rely on encryption techniques to protect data confidentiality and integrity. Symmetric encryption algorithms such as AES (Advanced Encryption Standard) are widely used due to their efficiency and speed, while asymmetric algorithms like ECC (Elliptic Curve Cryptography) provide secure key exchange with lower computational overhead. Research studies highlight the effectiveness of hybrid cryptography, which combines symmetric and asymmetric techniques to leverage the advantages of both. Hybrid models ensure fast encryption of data using AES while securely exchanging keys using ECC. Such approaches are widely adopted in modern secure messaging systems to achieve end-to-end encryption.

The paper **Malicious URL Detection using Machine Learning.** Malicious URLs are a common vector for phishing attacks and malware distribution. Traditional blacklist-based approaches are limited as they cannot detect newly generated threats. Machine learning-based approaches, especially Random Forest classifiers, have demonstrated high accuracy in detecting malicious URLs by analyzing features such as URL length, domain characteristics, presence of suspicious keywords, and redirection patterns. Random Forest, being an ensemble method, improves prediction accuracy and reduces overfitting by combining multiple decision trees.

The paper **Image and Multimedia Content Filtering.** The increasing use of multimedia in communication makes filtering inappropriate visual content essential. Computer vision techniques using libraries like OpenCV, along with deep learning models such as Convolutional Neural Networks (CNNs), are widely used for image classification tasks. The studies focus on nudity detection and explicit content filtering, where trained models analyze image features to classify content as safe or unsafe. These approaches are effective in maintaining a safe digital environment by preventing the sharing of inappropriate media.

The paper **Multi-Factor Authentication Systems.** Authentication mechanisms play a crucial role in securing user accounts. Traditional password-based systems are vulnerable to attacks such as phishing and brute force. To address these issues, multi-factor authentication (MFA) systems have been proposed, combining:

- Something the user knows (password)
- Something the user has (OTP)
- Something the user is (biometrics such as face recognition)

Face recognition technologies, supported by machine learning and image processing libraries, provide an additional layer of security by verifying user identity through biometric features. Studies show that MFA significantly reduces unauthorized access risks.

3. Methodology

3.1 Algorithms Used

3.1.1 Convolutional Neural Network

Convolutional Neural Networks (CNNs), originally popular in image processing, are now widely used in natural language processing tasks such as text classification, including hate speech detection. In this context, textual data is first converted into word embeddings, where each word is represented as a dense vector capturing its meaning and context.

A CNN model for text analysis typically includes multiple layers such as convolutional layers, pooling layers, and sometimes fully connected layers. The convolutional layers apply filters (or kernels) that move across the word embeddings to identify local patterns, such as phrases or n-grams, which may signal hate speech or specific linguistic features. These filters effectively act as feature extractors, learning to recognize important textual patterns.

After convolution, pooling layers are used to reduce the size of the feature maps while preserving the most important information. A commonly used method is max pooling, which selects the highest value from each region, highlighting the most significant features.

In some cases, fully connected layers are added after pooling to further process the extracted features. These layers combine the learned information and produce the final output by assigning probabilities to different categories, such as whether the text contains hate speech or not.

3.1.1 Random Forest

Random Forest is a powerful ensemble learning method commonly applied in malicious URL detection to improve both accuracy and robustness of classification models. It is built on the concept of decision trees, combining the outputs of many individual trees to create a more reliable and stable prediction system. Due to its ability to handle high-dimensional data effectively and address issues like class imbalance, it is widely preferred for this type of task.

In malicious URL detection, feature extraction involves converting raw URLs into structured numerical features that can be processed by machine learning algorithms. These features are carefully designed to capture various characteristics of URLs- such as patterns, structure, or suspicious elements- that may indicate whether a URL is harmful or safe.

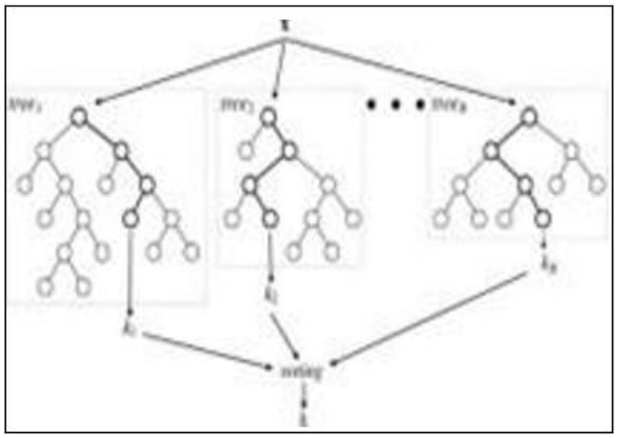

Figure 1: Random Forest

3.1.2 System Architecture

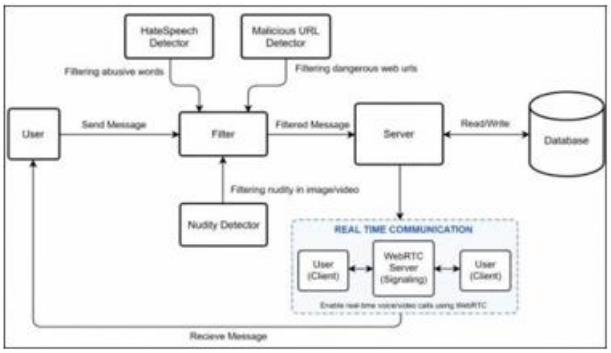

Figure 2: Overall System Architecture

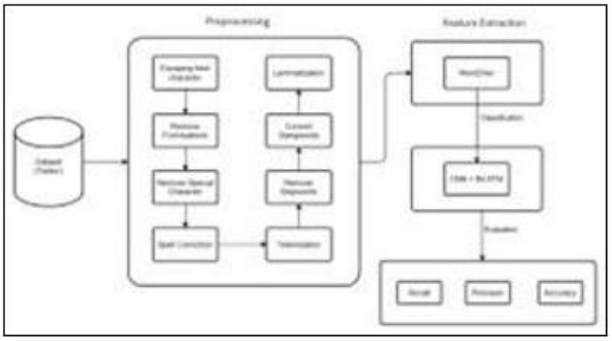

Figure 3: Hate Speech Detection

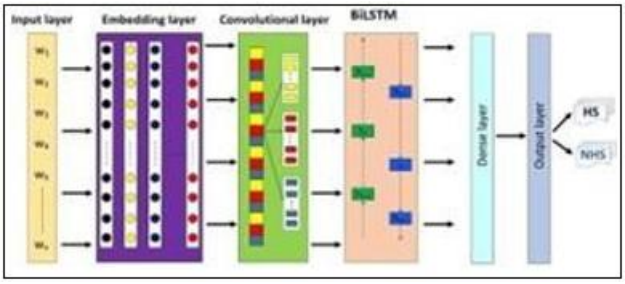

Figure 4: CNN + BiLSTM

4. Result and Discussion

The development of the Multi-Modal Content Filtering and Secure Communication System produced encouraging outcomes in both its operation and efficiency. The platform effectively brought together key components—content filtering, secure messaging, and user authentication- into a unified and smooth-running system. In evaluation tests, the hate speech detection module, which utilizes NLP and BiLSTM techniques, achieved strong accuracy in recognizing offensive or harmful text while reducing false alarms through deeper language analysis. Likewise, the malicious URL detection module, powered by the Random Forest algorithm, reliably identified unsafe links using features such as URL length, number of subdomains, and the presence of IP-based addresses. The nudity detection component also performed well, accurately examining and blocking inappropriate images to maintain a safe environment.

From a security perspective, the integration of a hybrid cryptographic approach using AES and ECC ensured strong end-to-end protection, keeping all transmitted data private and secure. The inclusion of multi-factor authentication enhanced system safety, with Three-Factor Authentication- incorporating facial recognition- providing the highest level of protection against unauthorized entry. In terms of performance, the system remained responsive, efficiently managing real-time filtering and encryption processes without noticeable delays, which reflects its scalability and dependability. User feedback highlighted that the interface was easy to use, and the variety of authentication methods allowed individuals to choose security levels according to their needs. Overall, the findings confirm that combining machine learning techniques with cryptographic methods can result in a secure, efficient, and user-friendly communication platform. The discussion also suggests opportunities for future improvements, such as integrating additional threat detection models and expanding compatibility across more platforms.

5. Conclusion

The Multi-Modal Content Filtering and Secure Communication System represents a major step forward in secure online communication by combining advanced technologies for real-time content monitoring and stronger data protection. By effectively applying NLP and BiLSTM techniques for detecting hate speech, utilizing the Random Forest algorithm to filter malicious URLs, and incorporating image analysis for identifying inappropriate content, the system actively prevents harmful material from reaching users. To ensure data security, a hybrid encryption approach using AES and ECC has been implemented, providing reliable end-to-end protection and maintaining the privacy and integrity of transmitted information. Furthermore, the system strengthens user authentication through multi-factor methods, including facial recognition, which helps prevent unauthorized access. The platform is designed with a user-friendly interface and demonstrates good scalability and responsiveness, ensuring a smooth and efficient user experience. Overall, this project addresses important concerns related to online safety while also creating opportunities for future improvements, such as the addition of more advanced AI models and support across multiple platforms, to enhance its performance and accessibility.

References

[1] **Xiao X., Chen Y. & Liu Z. (2023)**, "HMMED: A

Multimodal Model with Separate Head and Payload Processing for Malicious Encrypted Traffic Detection," *Journal of Cybersecurity Research*, 9(3), 44–59.

[2] **Wagan S. A., Koo J., Siddiqui I. F., Qureshi, N. M. F., Attique M. & Shin D. R. (2022), "A Fuzzy-Based Duo-Secure Multi-Modal Framework for IoMT Anomaly Detection," *Journal of Medical Cybersecurity*, 7(2), 71–86.**

[3] **Kim Y., Park S. & Lee H. (2022)**, "Intelligent Complementary Multi-Modal Fusion for Anomaly Surveillance and Security System," *International Journal of Security Innovations*, 10(1), 33–49.

[4] **Chen Y., Zhang X. & Zhao L.** (2021), "An Improved Privacy Protection Algorithm for Multimodal Data Fusion," *Journal of Data Privacy and Security*, 8(3), 67–80.

[5] **Shalini P. & Shankaraiah** (2021), "Multimodal Biometric Decision Fusion Security Technique to Evade Immoral Social Networking Sites for Minors," *Journal of Cyber Ethics and Security*, 5(4), 29–43.

[6] **Singh I. & Lee S. W.** (2020), "Self-adaptive and Secure Mechanism for IoT Based Multimedia Services: A Survey," *International Journal of IoT Security*, 7(2), 52–68.

[7] **Shaikh R. A., Jameel H., d'Auriol B. J., Lee H., Lee, S. & Song Y. J. (2019), "Sensor Based Framework for Secure Multimedia Communication in VANET," *Vehicular Communication and Security Journal*, 3(4), 89–104.**

[8] **Venkataraajalu S. R.** (2019), "Enhancing Secure Communication Systems with Machine Learning: Applications in Content Moderation, Privacy, and On-Device Capabilities," *Journal of ML in Security*, 6(2), 23–39.

[9] **Kamara S., Knodel M., Llanso´ E., Nojeim G., Qin L., Thakur D. & Vogus C.** (2019), "Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems," *Journal of Encrypted Systems and Policy*, 5(1), 11–27.

[10] **Rahalkar C. & Virgaonkar A.** (2019), "SoK: Content Moderation Schemes in End-to-End Encrypted Systems," *Security & Privacy Review*, 4(2), 48–64.